# Long-term musical experience and auditory and visual perceptual abilities under adverse conditions

Esperanza M. Anaya[a)]
*Department of Behavioral Science, Midwestern University, 555 31st Street, Downers Grove, Illinois 60515, USA*

David B. Pisoni[b)]
*Department of Psychological and Brain Sciences, Indiana University, 1101 East 10th Street, Bloomington, Indiana 47405, USA*

William G. Kronenberger[c)]
*Department of Otolaryngology–HNS, Indiana University School of Medicine, 705 Riley Hospital Drive, Indianapolis, Indiana 46202, USA*

(Received 23 September 2015; revised 27 July 2016; accepted 18 August 2016; published online 28 September 2016)

Musicians have been shown to have enhanced speech perception in noise skills. It is unclear whether these improvements are limited to the auditory modality, as no research has examined musicians' visual perceptual abilities under degraded conditions. The current study examined associations between long-term musical experience and visual perception under noisy or degraded conditions. The performance of 11 musicians and 11 age-matched nonmusicians was compared on several auditory and visual perceptions in noise measures. Auditory perception tests included speech-in-noise tests and an environmental sound in noise test. Visual perception tasks included a fragmented sentences task, an object recognition task, and a lip-reading measure. Participants' vocabulary knowledge and nonverbal reasoning abilities were also assessed. Musicians outperformed nonmusicians on the speech perception in noise measures as well as the visual fragmented sentences task. Musicians also displayed better vocabulary knowledge in comparison to nonmusicians. Associations were found between perception of speech and visually degraded text. The findings show that long-term musical experience is associated with modality-general improvements in perceptual abilities. Possible systems supporting musicians' perceptual abilities are discussed.
 [http://dx.doi.org/10.1121/1.4962628]

[MSS] Pages: 2074–2081

## I. INTRODUCTION

In everyday listening environments, individuals are typically confronted with multiple acoustic signals and are challenged with the task of focusing on and identifying specific aspects of the signal. An example of this is the problem of recognizing speech in noise or under other adverse conditions such as competing talkers, reverberation, and masking from environmental sounds (Mattys *et al.*, 2012). Musicians have been shown to have enhanced speech perception skills in adverse listening conditions when compared to individuals with no prior musical experience (Fuller *et al.*, 2014; Parbery-Clark *et al.*, 2009; Parbery-Clark *et al.*, 2011; Soncini and Costa, 2006; Strait and Kraus, 2011a). A growing body of research has suggested close connections between music and language (Koelsch *et al.*, 2005; Moreno, 2009; Patel, 2008). Several studies that examined the relations between musical training and language abilities have found that musical abilities are associated with improved vocabulary (Forgeard *et al.*, 2008) as well as enhanced subcortical processing for syllables presented in quiet listening conditions (Musacchia *et al.*,2008).

While it is still unclear what neural pathways and cognitive factors are associated with the effects of long-term musical experience, Parbery-Clark *et al.* (2011) and Strait and colleagues (Strait and Kraus, 2011b) have argued that musicians' superior speech perception skills stem from strengthened auditory executive functions. This stands in contrast to findings from nonmusician literature, which has shown links between the ability to segregate and selectively focus conscious attention on a target speech signal in noise to modality general cognitive abilities related to attentional and inhibitory control (Heinrich *et al.*, 2008; Obleser *et al.*, 2007; Pichora-Fuller *et al.*, 1995; Rabbitt, 1968). Parbery-Clark *et al.* (2011) and Strait and Kraus (2011a; Strait *et al.*, 2012) reported that musical training is associated with enhanced speech perception in noise skills as well as improved executive function skills; yet, these enhanced executive function skills were auditory specific and were shown in verbal working memory and auditory attention, while no group difference was shown for visual-spatial cognitive abilities between musicians and nonmusicians. These researchers suggest that musicians are a unique population with long-term experience processing complex auditory signals and that these unique auditory experiences have led to

[a)]Electronic mail: eanaya@midwestern.edu
[b)]Also at: Department of Otolaryngology–HNS, Indiana University School of Medicine, Indianapolis, IN 46202, USA.
[c)]Also at: Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN 46202, USA.

a specific strengthening of auditory cognitive abilities in comparison to visual cognitive abilities.

Musicians' improved speech perception in noise skills may be supported by their improvements in executive function and cognitive control (Kraus *et al.*, 2012; Parbery-Clark *et al.*, 2011; Strait and Kraus, 2011a; Strait *et al.*, 2012; Zuk *et al.*, 2014). Yet, this claim that musicians have auditory specific perceptual enhancements still remains to be determined. Within the literature that has examined musicians' auditory perceptual abilities under adverse listening conditions, there is a discrepancy with some researchers finding that musicians have enhanced speech perception abilities relative to nonmuscians (Parbery-Clark *et al.*, 2009; Parbery-Clark *et al.*, 2011; Soncini and Costa, 2006; Strait *et al.*, 2010) and other research showing no group differences on these measures (Boebinger *et al.*, 2015; Fuller *et al.*, 2014; Zendel and Alain, 2012). These discrepancies may stem from methodological differences between studies, such as noise used to degrade the speech signal (Boebinger *et al.*, 2015; Fuller *et al.*, 2014) or the definition of musician (Zendel and Alain, 2012). While there is a growing body of literature aiming to better understand musicians' auditory perception under adverse conditions, there has been no research carried out on musicians' perception of degraded visual information.

The argument for selective strengthening of auditory cognitive abilities over visual cognitive abilities stands in contrast to previous research linking visual and auditory processes (Watson *et al.*, 1996). Previous research suggests that an enhancement in auditory perceptual abilities may also be associated with enhanced visual perceptual abilities (Besser *et al.*, 2012; Krull *et al.*, 2012; Watson *et al.*, 1996). Watson *et al.* (1996) found an association between nonmusician adults' ability to recognize degraded auditory speech and their ability to recognize degraded visual speech and text (e.g., lip-reading and reading degraded sentences). Watson and colleagues argue that auditory and visual speech perception are dependent on the same general information processing system. More recent work provides additional converging support for the major theoretical claim by Watson *et al.* that auditory and visual perceptual abilities are closely linked (Besser *et al.*, 2012; Krull *et al.*, 2012). It has been suggested that executive functions underlie both speech perception abilities as well as visual perceptual abilities (Feld and Sommers, 2009; Watson *et al.*, 1996). If speech recognition is mediated by a domain-general processing system, then one would predict that musicians who show enhancements in speech recognition in noisy conditions would also show enhancements in the perception of degraded visual information.

In summary, there are two competing hypotheses regarding the benefits of long-term musical experience: a modality specific hypothesis (only auditory experience and processing shows benefit) and a modality general hypothesis (the benefit goes beyond processing auditory information and may also affect visually presented verbal information or nonverbal visual information). The current study was designed to elucidate the modality-specific vs modality-general nature of musicians' perceptual skills under adverse information processing conditions, specifically, we examined musicians' auditory and visual perception skills in noise. We hypothesized a modality-general effect showing that musicians would exhibit better perceptual skills and perform better on both auditory and visual perceptual tasks under degraded presentation conditions when compared to nonmusicians.

## II. METHODS

### A. Participants

Twenty-two age matched ($M = 20.72$, $SD = 2.72$) individuals participated in this study. All participants completed the informed consent that was approved by Indiana University's Internal Review Board and all participants were monolingual speakers of American English. Eleven participants were identified as musicians and were recruited from Indiana University's Jacob School of Music. All musicians played either piano or organ and began their musical training at or before the age of nine ($M = 4.9$, $SD = 1.44$), had on average 15.45 yr of musical training ($SD = 2.69$), and continued to practice their instrument regularly (hours practice per week, $M = 16.31$, $SD = 10.44$). Eleven nonmusician participants were recruited. This group had little-to-no experience playing an instrument (experience playing in years, $M = 1.72$, $SD = 1.9$). All nonmusicians reported no longer playing any musical instrument at time of testing. Nonmusicians were recruited through the use of flyers that were posted around the Indiana University Bloomington campus.

### B. Materials and procedures

All tasks, with the exception of matrix reasoning, were conducted in a sound-attenuated IAC booth in the Speech Research Laboratory at Indiana University Bloomington. For the auditory tasks, stimuli were presented over high-quality headphones (Beyerdynamic DT109). All participants initially completed a pure-tone hearing test. All participants exhibited normal hearing [$\leq$20 dB hearing level (HL) pure tone thresholds from 250 to 8000 Hz].

### C. Measures

#### *1. Matrix reasoning*

The matrix reasoning subtest of the WASI II (Wechsler, 2011) was administered to obtain a normed baseline measure of global nonverbal intelligence. Matrix reasoning is used to assess nonverbal abstract problem solving abilities. Participants were shown an array of visual images with one missing square and were required to fill in the missing portion of the abstract patterns by selecting an image that best fit the array from five picture options. The task was terminated when participants were unable to identify the correct pattern in four consecutive trials. A *t*-score was calculated for each participant based on his/her raw score (the number of correctly completed patterns).

#### *2. Boston naming*

The Boston naming test was used to assess participants' word retrieval skills (Kaplan *et al.*, 2001). Participants were shown an image of a common object (e.g., beaver or canoe) and were asked to identify the object as quickly and as accurately as possible. Sixty images were shown individually on a computer monitor. Participants were given 20 s to correctly name the object before the task advanced to the next trial. An overall percent correct recognition score and response latencies were obtained for each participant.

### 3. Auditory perception tests

*a. Hearing-in-noise-test (HINT) and perceptually robust English sentence test (open-set) (PRESTO).* The HINT was developed by Nilsson *et al.* (1994) to measure sentence recognition skills in noise. In this task, participants were presented with a sentence and were asked to repeat back what they heard. Sentences were presented in speech-shaped noise that matched the long-term spectrum of the entire set of stimulus sentences. The sentences used on the HINT originally came from materials developed by Bench *et al.* (1979). The HINT test contains 25 sentence lists with 10 sentences in each list. Participants were presented with HINT lists 1–5 for a total of 50 sentences. There were a total of 168 target words. The sentences were presented at one signal-to-noise ratio of −3 dB SNR. Previous research has shown that this SNR is a challenging presentation condition that yields a wide range of scores (Parbery-Clark *et al.*, 2009).

The PRESTO test was developed by Gilbert *et al.* (2013). Each sentence in a list was spoken by a different talker and none of the talkers were repeated within a list. Participants were presented with a spoken sentence and repeated back what they heard. Participants were presented with a total of 50 PRESTO sentences that were mixed with six-talker babble at 0 dB SNR and contained between five and 10 words. There were a total of 208 target words in the task.

For the HINT and PRESTO tasks, participants' responses were scored by calculating and totaling the number of key words correctly identified for each sentence. HINT and PRESTO scores were then transformed into standardized z scores, combined and averaged to create a speech perception in noise (SIN) composite score.

*b. Environmental sounds test.* An environmental sounds test was administered to assess participants' ability to recognize common environmental sounds under degraded listening conditions. Participants were presented with a nonspeech environmental sound that was mixed with white noise at −5 dB SNR and were asked to identify the sound. The environmental sounds task used an open-set response format. Participants heard a total of 25 environmental sounds. Stimuli were taken from the environmental sound database originally created by Marcell *et al.* (2000). The database contains 120 sounds from various categories: animals, vehicles, nonspeech sounds made by humans, music, and others (glass breaking, police siren, etc.). Stimuli have been normed using a group of typically developing adults on several indexes including familiarity and naming response latency (Marcell *et al.*, 2000). The environmental sounds are available on the internet (http://ww.cofc.edu/∼marcellm/confront.htm). Musical sounds were not used as stimuli because it was assumed musicians would show a response bias in correctly identifying musical instruments in comparison to nonmusicians. The white noise was created using Audacity software. A percent correct score was computed for each participant by calculating the number of environmental sounds correctly identified.

### 4. Visual perception tests

*a. Fragmented sentence test.* The fragmented sentence test developed by Feld and Sommers (2009) based on earlier work by Watson *et al.* (1996) was used to assess visual perceptual abilities for recognizing printed words in degraded sentences. In this task, participants were shown a visually degraded meaningful English sentence on a computer screen and were required to read the sentence aloud to the experimenter. Figure 1 (Panel B) shows an example of the visual display of two degraded sentences ("The garden needs to be replanted next spring." and "He thinks it's warm enough to go to the beach."). Participants viewed 35 visually presented sentences. Stimuli consisted of CUNY sentences that ranged in length from five to 12 words (Boothroyd *et al.*, 1988). Sentences also varied in semantic content and contained either low or moderately predictable contextual information. Sentence materials were provided to us by Dr. Mitchell Sommers of Washington University (Feld *et al.*, 2009). Stimuli were presented using Arial font on a computer screen for 3 s. Sentences were degraded by deleting 60 random pixels from each letter. There were a total of 178 key words. Participants' responses were scored by recording the number of key words they were able to correctly identify from each display.

*b. Object recognition.* An object recognition task was used to assess participants' visual perception for objects presented with visual noise. Participants were shown familiar visual objects that were masked with visual noise and were

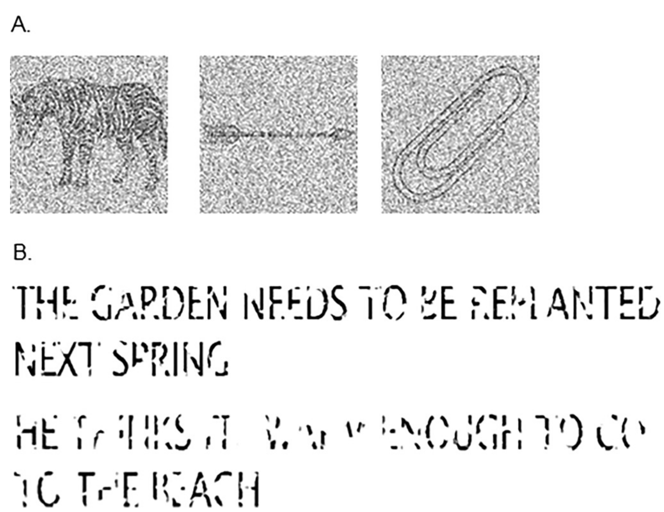


FIG. 1. Sample stimuli for the visual perception in noise tasks. Sample stimuli for the object recognition task are displayed in panel A (zebra, arrow, paperclip). Sample stimuli for the Fragmented Sentences Task are shown in panel B ("The garden needs to be replanted next spring"; "He thinks it's warm enough to go to the beach").

asked to identify the objects. Figure 1 (panel A) shows several sample objects (zebra, arrow, paperclip). Images were shown individually on the computer screen for 250 ms. After an object was presented, participants reported their responses aloud. Participants pressed the space bar to advance to the next trial. Images were taken from the International Picture Naming Project, University of California, San Diego (Szekely *et al.*, 2004). A total of 35 visual images were selected. Images consisted of black and white line drawings. All stimuli were high frequency images that measured at $300 \times 300$ pixels. A white noise was created to mask the stimulus images. The white noise was created in Adobe Photoshop using the filter function. A filter was overlaid onto each object and set at 100% monochromatic. Each filter was then dissolved down to 30%. Responses were scored as percent correct object recognition.

*c. Lip-reading.* A measure of lip-reading was also obtained to assess participants' perception of visual speech information. In this task, participants were presented with video recordings of CUNY sentences (Boothroyd *et al.*, 1988) where the audio tracks were removed. Participants saw one speaker say, a sentence and they were required to report aloud what words they thought the speaker said. Participants were presented with a total of 20 CUNY sentences, at lengths of 3, 5, 7, 9, and 11 words. Participants completed four sentences at each length. The female talker spoke in a slow, well-articulated manner. The speaker also exhibited neutral speech prosody and facial emotions. There were a total of 140 key words. A percent correct score was computed for each participant by calculating the number of key words correctly identified across all 20 sentences.

## III. RESULTS

### A. Preliminary analyses

Preliminary ANOVAs were conducted in order to examine group differences between musicians and nonmusicians in years of education, vocabulary and nonverbal reasoning. No group differences were found for years of education ($p = 0.748$) or nonverbal reasoning ($p = 0.301$). A significant group difference was found for Boston naming, $F(1, 20) = 5.15$, $p = 0.034$. Musicians exhibited better vocabulary knowledge than nonmusicians. Descriptive statistics for matrix reasoning and Boston are listed in Table I.

Additional analyses were carried out on the Boston naming task to examine participants' naming speed. Audio recordings were made of each participant naming the stimulus objects. Response latencies were taken from the onset of the visual stimulus to the onset of the participant's vocal response. Only correct responses were measured. Response latencies were analyzed for the first 40 images that were shown. This was done in order to examine naming speed without confounding effects of group differences in accuracy of identifying the objects. Results showed no differences in naming accuracy for the first 40 items, $F(1, 20) = 1.4$, $p = 0.25$, between groups. However, using the first 40 trials, group differences were found in naming speed, $F(1, 20) = 7.79$, $p = 0.011$, with musicians exhibiting faster naming speeds over nonmusicians.

TABLE I. Musicians vs nonmusicians group results as shown by univariate analyses. Note: M: Mean; SD: Standard deviation. All scores are percent correct except for the SIN composite (averaged *z*-score), matrix reasoning (*t*-score), Boston naming accuracy (raw score), and Boston RL (response latencies in seconds).

| | Musicians | | Nonmusicians | | | |
|---|---|---|---|---|---|---|
| **Outcome measures** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***F*(20)** | ***p*** |
| **SIN composite** | 0.32 | 0.81 | −0.39 | 0.87 | 4.07 | 0.057 |
| **Environmental sounds** | 67.64 | 17.01 | 70.55 | 8.44 | 0.25 | 0.617 |
| **Fragmented sentences** | 83.25 | 7.27 | 72.83 | 8.74 | 9.22 | 0.007 |
| **Object recognition** | 65.45 | 9.51 | 58.96 | 10.24 | 2.37 | 0.139 |
| **Lip-reading** | 5.26 | 3.56 | 6.23 | 4.81 | 0.291 | 0.595 |
| **Matrix reasoning** | 55.45 | 2.91 | 51.45 | 12.14 | 1.12 | 0.301 |
| **Boston naming** | 56.54 | 1.69 | 52.72 | 5.31 | 5.15 | 0.034 |
| **+Boston RL (s)** | 1.62 | 0.14 | 1.86 | 0.24 | 7.79 | 0.011 |

### B. Perception in noise tasks

A MANOVA was used to examine group differences for the measures of auditory and visual perception in adverse conditions. Box's M test was not significant ($p = 0.397$). A significant group difference was found for the perception tests, Wilks' Lambda $= 3.04$, $F(1,16) = 1.98$, $p = 0.041$. Subsequent univariate analyses revealed a marginal group difference for the SIN composite score, $F(1,20) = 4.07$, $p = 0.057$, $\eta^2 = 0.169$. The environmental sounds test violated the homogeneity assumption as assessed by the Levene test $F(1,20) = 6.39$, $p = 0.022$. To evaluate this effect, we used a more stringent alpha level of 0.01. No group difference was found for environmental sounds, $p = 0.617$. For the visual perception in noise tests, a group difference was found for fragmented sentences, $F(1,20) = 9.22$, $p = 0.007$, $\eta^2 = 0.316$, with musicians performing better than nonmusicians. No group differences were found for degraded visual object recognition, $p = 0.139$, or lip-reading, $p = 0.595$. Overall, a marginally significant group difference was found for speech perception in noise (SIN composite) and a significant group difference was found for the visual perception of degraded text (fragmented sentences test) with musicians outperforming nonmusicians.

### C. Relations among measures

#### 1. Groups collapsed

Bivariate correlations were carried out to assess relations among the measures. A correlation matrix of these results with groups collapsed is shown in Table II. Associations were found between speech perception and visual perception in noise measures. The SIN composite scores were correlated with fragmented sentence scores ($r = 0.646$, $p = 0.001$; see Fig. 2 for a scatterplot that displays the relations between participants' performance on these two measures). Object recognition scores were also moderately correlated with SIN composite scores ($r = 0.388$, $p = 0.075$) and environmental

TABLE II. Correlations between behavioral measures and demographic information (groups collapsed). YrPl: Years of playing a musical instrument; Mat: matrix reasoning; Bost; Boston naming accuracy; BoRL: Boston response latency; SIN: speech perception in noise composite; Env: environmental sounds; FragS: fragmented sentences; ObjR: object recognition; LipR: lip-reading.

| | YrPl | Mat | Bost | BoRL | SIN | Env | FragS | ObjR | LipR |
|---|---|---|---|---|---|---|---|---|---|
| YrPl | - | - | - | - | - | - | - | - | - |
| Mat | 0.233 | - | - | - | - | - | - | - | - |
| Bost | 0.419 | 0.482[a] | - | - | - | - | - | - | - |
| BoRL | −0.551[b] | −0.491[a] | −0.651[b] | - | - | - | - | - | - |
| SIN | 0.471[a] | 0.208 | 0.761[c] | −0.504[a] | - | - | - | - | - |
| Env | −0.048 | 0.103 | 0.098 | −0.05 | 0.314 | - | - | - | - |
| FragS | 0.612[b] | 0.457[a] | 0.481[a] | −0.456[a] | 0.641[b] | 0.122 | - | - | - |
| ObjR | 0.379 | 0.332 | 0.415 | −0.422 | 0.388 | 0.491[a] | 0.593[b] | - | - |
| LipR | −0.104 | 0.342 | 0.02 | −0.16 | 0.185 | −0.03 | 0.349 | −0.1 | - |

[a] $p < 0.05$
[b] $p < 0.01$
[c] $p < 0.001$ (two-tailed).

sound scores ($r = 0.491$, $p = 0.02$). Correlations found between auditory perception and visual perception in noise tasks suggest that these tasks are measuring the same underlying information processing skills.

Associations were also found between perception in noise measures and verbal skills as well as nonverbal reasoning abilities. SIN composite scores were correlated with Boston naming accuracy ($r = 0.761$, $p < 0.001$) and Boston response latencies (naming speed) ($r = -0.504$, $p = 0.017$). Performance on the fragmented sentences test was also correlated with Boston naming accuracy ($r = 0.481$, $p = 0.023$) and Boston naming speed ($r = -0.456$, $p = 0.033$). These associations suggest that individuals who had larger vocabularies and faster naming speeds exhibited better speech-perception-in-noise skills as well as visual perception of degraded text than individuals who had smaller vocabularies and slower processing speeds. Furthermore, performance on the matrix reasoning task correlated with fragmented sentences ($r = 0.457$, $p = 0.032$) but not SIN composite ($r = 0.208$, $p = 0.353$). Years of musical training was correlated with SIN composite ($r = 0.471$, $p = 0.027$) and fragmented sentences ($r = 0.612$, $p = 0.002$) as well as with Boston response latencies ($r = -0.551$, $p = 0.008$), and moderately correlated with Boston naming accuracy ($r = 0.419$, $p = 0.052$).

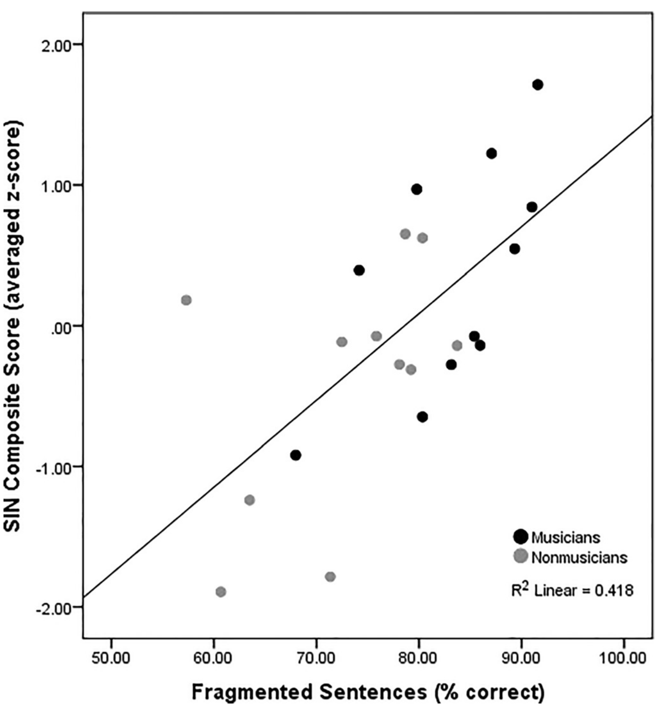


FIG. 2. Relations between speech perception in noise skills and visual perception in noise skills as seen in participants' SIN performance and Fragmented Sentence tasks. Participants' SIN performance is displayed on the $y$-axis (averaged $z$-score) and participants' performance on the fragmented sentence task is shown on the $x$ axis (% correct). Musicians are displayed in black and nonmusicians in grey. Better speech perception in noise skills was associated with better visual perception under adverse conditions.

### 2. Groups separated

Bivariate correlations were also carried out for each individual group. For musicians, the SIN composite scores were correlated with fragmented sentence scores, $r = 0.626$, $p = 0.04$; no significant association was found for nonmusicians, $r = 0.494$, $p = 0.122$. For nonmusicians, object recognition scores correlated with environmental sound scores ($r = 0.76$, $p = 0.007$); no significant association was found for musicians, $r = 0.515$, $p = 0.105$. Years of musical training correlated with SIN composite ($r = 0.699$, $p = 0.017$) for the musicians. While not statistically significant, some relation was shown between years of musical training and scores on fragmented sentences ($r = 0.521$, $p = 0.1$). For nonmusicians, no associations were found between years of musical training and SIN scores ($p = 0.453$) or fragmented sentences ($p = 0.829$). It should be noted that power may have been affected in the separation of these groups, as can be seen in the moderate to strong correlations between measures where no statistically significant association was found (e.g., association between SIN composite and fragmented sentences for nonmusicians).

## D. Effects of vocabulary

The group difference found between musicians and nonmusicians on the auditory and visual perceptual measures could be the result of differences in vocabulary. As stated earlier, there was a significant difference in vocabulary knowledge between groups with musicians exhibiting a larger vocabulary than nonmusicians. It is possible that musicians' enhanced verbal perceptual abilities could be attributed to their improved vocabulary knowledge. In order to overcome this potential confounding variable, we conducted additional univariate analyses to assess the influence of vocabulary knowledge on speech perception and degraded text perception.

An ANCOVA was conducted with SIN composite as the dependent variable and Boston scores entered as a covariate. No group difference was found in speech perception after controlling for vocabulary knowledge, $p = 0.619$. Another ANCOVA was conducted with fragmented sentences entered as the dependent variable and Boston scores again entered as the covariate. Even after controlling for differences in vocabulary, a group difference was found between musicians and nonmusicians on the fragmented sentences task, $F(1,19) = 4.56$, $p = 0.046$. These results suggest that differences in vocabulary knowledge cannot account for the performance differences seen between musicians and nonmusicians on the fragmented sentences task.

## IV. DISCUSSION

The present study was conducted to investigate the modality-specific nature of musicians' perceptual information processing skills under degraded presentation conditions. We hypothesized that long-term formal musical training and experience would be associated with modality-general enhancements in several perceptual abilities. We predicted that musicians would exhibit enhanced perceptual skills and perform better on both auditory and visual perception tasks under adverse/degraded conditions.

The results obtained in this study using a range of visual and auditory information processing tasks suggest that musicians display enhancements that are not limited exclusively to auditory perceptual processing skills. A marginally significant group difference with a moderate effect size was found for speech perception in noise. These results support the findings of previously conducted research that showed associations between long-term musical training and enhanced speech perception in noise (Parbery-Clark *et al.*, 2009; Parbery-Clark *et al.*, 2011; Soncini and Costa, 2006; Strait *et al.*, 2010). Of particular interest was the robust group difference, with a large effect size, that was found in performance on the fragmented sentences task with musicians out performing nonmusicians. Overall, musicians exhibit enhanced performance for verbal processing under challenging auditory or visual conditions since the SIN measures and the fragmented sentences task explicitly rely on verbal information processing skills. This is the first study to show that long-term formal musical experience is associated with enhanced perception of degraded visual text.

Associations were also found between speech perception and visual perception of text, as shown by correlations between SIN composite scores and performance on the fragmented sentences test. These findings provide additional converging support for earlier reports suggesting links between auditory and visual perceptual abilities under degraded presentation conditions (Krull *et al.*, 2012; Watson *et al.*, 1996). Relations between auditory and visual perceptual abilities have been examined using the text reception threshold test, a visual analogue of the speech reception threshold test (Zekveld *et al.*, 2007). In the text reception threshold test, participants are presented with varying degrees of masked text and they are asked to identify the text. A threshold is identified where participants can accurately identify 50% of the text. Several studies have found associations between performance on the text reception threshold test and performance on the speech reception threshold test (Besser *et al.*, 2013; Besser *et al.*, 2012; George *et al.*, 2007; Zekveld *et al.*, 2007). These findings suggest that speech perception in either the auditory or visual modalities rely on a shared information processing system.

Watson and colleagues (1996) have argued that this shared, underlying system is comprised of domain-general processes. Executive functions are a domain-general system that support processes in other cognitive systems (Miyake and Friedman, 2012), such as language (Kronenberger *et al.*, 2014; Luria, 1973). The term executive function is an umbrella term used to encompass functions such as attention, inhibition, emotional control, working memory, initiation, and the ability to plan, organize, and prioritize actions and activities. Specifically, the executive function subcomponent of working memory has been shown to support speech perception abilities under adverse conditions (Francis and Nusbaum, 2009; Koelewijn *et al.*, 2012; Mattys *et al.*, 2012; Pichora-Fuller *et al.*, 1995; Rönnberg *et al.*, 2010). Musicians have been shown to have enhanced working memory skills (George and Coch, 2011; Parbery-Clark *et al.*, 2011; Strait *et al.*, 2012). Zuk and colleagues (2014) argue that musical training may strengthen the domain-general system of executive functions, which may then support and strengthen other cognitive abilities. In regards to the findings from the current study, musicians' enhanced perceptual skills for challenging conditions may stem from their improved executive functions, specifically, their working memory abilities.

In this study, we replicated earlier studies that showed links between musical experience and language abilities. The duration of musical training, as measured by the number of years of musical training, correlated significantly with several speech and language abilities. Individuals with more musical experience had better SIN composite scores, better performance on the fragmented sentences test, and had faster Boston picture naming speeds. Several researchers have argued that there are associations between music and language processes (Koelsch *et al.*, 2005; Moreno, 2009; Moreno *et al.*, 2011; Patel, 2008). Most notably, both music and spoken language make use of dynamic time-varying spectrotemporal information encoded in auditory signals. Pitch information is integral in music perception whereas pitch is used to convey prosodic as well as indexical information in speech. Music and language also consist of hierarchical structures and complex temporal patterns that change over time. In music, harmonic and metric information are both highly structured in a hierarchical manner (Patel, 2008). The hierarchical structure of spoken language also consists of different levels, such as phonemes, syllables, words, and sentences. Semantic and syntactic information is also present in both systems (Koelsch *et al.*, 2005; Koelsch *et al.*, 2004; Patel *et al.*, 1998). Furthermore, music and language contain statistical regularities that can be implicitly learned starting at a young age (Saffran *et al.*, 1996; Saffran *et al.*, 1999). Given these similarities, it has been suggested that music and language may share a common neural biological processing system (Patel, 2011). While this shared information processing system remains to be described in greater detail, there is sufficient evidence reported in the literature to suggest that music and language processes overlap in some way and a strengthening of one system appears to affect processes in the other system (Bangert *et al.*, 2006; Chartrand and Belin, 2006; Gaab *et al.*, 2006; Jentschke and Koelsch, 2009; Moreno, 2009; Moreno *et al.*, 2011).

There were several limitations in the current study. A large limitation is that of our sample size, which consisted of eleven participants in each group. This small sample size affected statistical power, which could explain some of the correlational analyses that we found when groups were separated. An additional limitation can be seen in the noise that was selected for the environmental sounds task. The white noise that was used to mask the target stimulus may not have covered the long-term-spectra of each stimulus, which would have made the stimuli easier to identify. Furthermore, the nature of this study is correlational. We are unable to make statements regarding causation and can only state that there are associations between musical experiences and enhanced perceptual abilities.

## V. CONCLUSION

The objective of the study was to assess the modality-specificity of musicians' perceptual skills under adverse conditions. The present findings suggest that long-term formal musical experience is associated with modality-general enhancement of verbal perceptual abilities. Musicians outperformed nonmusicians on perceptually challenging behavioral tasks that included recognizing speech in noise, as well as recognizing visually degraded meaningful sentences. The present findings provide novel evidence showing that musicians' enhanced verbal perceptual abilities for degraded conditions extend beyond just the auditory modality. In addition, we found that musicians have better vocabulary knowledge, as well as faster naming speeds, in comparison to nonmusicians. Musicians' enhanced vocabulary knowledge may contribute to their perceptual skills when conditions are degraded. However, group differences on the fragmented sentences test remained after controlling for vocabulary knowledge. This finding, as well as results showing strong associations between auditory and visual perceptual tasks, suggests that perceptual abilities under adverse conditions may partially rely on shared domain-general information processing system, such as executive functions. Future research needs to examine in greater depth the role of language skills in relation to musicians' enhanced perceptual abilities. Other research should also assess if there are associations between musicians' perceptual abilities for degraded conditions and their executive functions. Furthermore, replicating earlier studies, we found that musicians performed better than nonmusicians in recognizing speech in noise; however, we failed to find any group difference in a task that involved identifying environmental sounds in noise. Additional studies should be carried out to explore whether musicians' superior perceptual processing skills are specific to language processing abilities regardless of input modality or whether they extend more broadly to other more general domains of cognitive functioning such as categorization of visual-spatial and temporal nonspeech patterns, memory, and learning.


## ACKNOWLEDGMENTS

This work is supported by NIH-NIDCD: 1R01 DC000064 and T32 DC00012.



Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H.-J., and Altenmuller, E. (**2006**). "Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction," Neuroimage **30**(3), 917–926.

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partial-hearing children," Br. J. Audiol. **13**(3), 108–112.

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., and Festen, J. M. (**2013**). "How linguistic closure and verbal working memory relate to speech recognition in noise—A review," Trends Amplif. **17**(2), 75–93.

Besser, J., Zekveld, A. A., Kramer, S. E., Ronnberg, J., and Festen, J. M. (**2012**). "New measures of masked text recognition in relation to speech-in-noise perception and their associations with age and cognitive abilities," J. Speech, Lang. Hear. Res. **55**, 194–209.

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., and Scott, S. K. (**2015**). "Musicians and non-musicians are equally adept at perceiving masked speech," J. Acoust. Soc. Am. **137**(1), 378–387.

Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (**1988**). "Voice fundamental frequency as an auditory supplement to the speech-reading of sentences," Ear Hear. **9**, 306–312 (1988).

Chartrand, J. P., and Belin, P. (**2006**). "Superior voice timbre processing in musicians," Neurosci. Lett. **405**, 164–167.

Feld, J. E., and Sommers, M. S. (**2009**). "Lipreading, processing speed, and working memory in younger and older adults," J. Speech Lang. Hear. Res. **52**, 1555–1565.

Forgeard, M., Winner, E., Norton, A., and Schlaug, G. (**2008**). "Practicing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning," PLoS One **3**(10) e3566.

Francis, A. L., and Nusbaum, H. C. (**2009**). "Effects of intelligibility on working memory demands for speech perception," Atten. Percept. Psychophys. **71**(6), 1360–1374.

Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., and Baskent, D. (**2014**). "The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations?" Front. Neurosci. **8**, 179.

Gaab, N., Tallal, P., Kim, H., Lakshminarayanan, K., Archie, J. J., Glover, H., and Gabrieli, J. D. E. (**2006**). "Neural correlates of rapid spectrotemporal processing in musicians and nonmusicians," Ann. N.Y. Acad. Sci. **1060**, 82–88.

George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (**2007**). "Auditory and nonauditory factors affecting speech reception in noise by older listeners," J. Acoust. Soc. Am. **121**(4), 2362–2375.

George, E. M., and Coch, D. (**2011**). "Music training and working memory: An ERP study," Neuropsychologia **49**(5), 1083–1094.

Gilbert, J. L., Tamati, T. N., and Pisoni, D. B. (**2013**). "Development, reliability, and validity, of PRESTO: A new high-variability sentence recognition test," J. Am. Acad. Audiol. **24**, 26–36.

Heinrich, A., Schneider, B. A., and Craik, F. I. M. (**2008**). "Investigating the influence of continuous babble on auditory short-term memory performance," Q. J. Exp. Psychol. **61**(5), 735–751.

Jentschke, S., and Koelsch, S. (**2009**). "Musical training modulates the development of syntax processing in children," Neuroimage **47**, 735–744.

Kaplan, E., Goodglass, H., and Weintraub, S. (**2001**). *Boston Naming Test Second Edition* (Pro-ed, Austin, TX).

Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., and Kramer, S. E. (**2012**). "Processing load induced by informational masking related to linguistic abilities," Int. J. Otolaryngol. **2012**, 865731.

Koelsch, S., Gunter, T., Wittfoth, M., and Sammler, D. (**2005**). "Interaction between syntax processing in language and in music: An ERP study," J. Cognit. Neurosci. **17**(10), 1565–1577.

Koelsch, S., Kasper, E., Sammler, D., Schulze, T. G., and Friederici, A. D. (**2004**). "Music, language and meaning: Brain signatures of semantic processing," Nat. Neurosci. **7**(3), 302–307.

Kraus, N., Strait, D., and Parbery-Clark, A. (**2012**). "Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory," Ann. N.Y. Acad. Sci. **1252**, 100–107.

Kronenberger, W. G., Colson, B. G., Henning, S. C., and Pisoni, D. B. (**2014**). "Executive functioning and speech-language skills following long-term use of cochlear implants," J. Deaf Stud. Deaf Educ. **19**(4), 456–470.

Krull, V., Humes, L. E., and Kidd, G. R. (**2012**). "Reconstructing wholes from parts: Effect of modality, age, and hearing loss, on word recognition," Ear Hear. **34**(2), 14–23.

Luria, A. R. (**1973**). *The Working Brain an Introduction to Neuropsychology* (Basic Books, New York).

Marcell, M. M., Borella, D., Greene, M., Kerr, E., and Rogers, S. (**2000**). "Confrontation naming of environmental sounds," J. Clin. Exp. Neuropsychol. **22**(6), 830–864.

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (**2012**). "Speech recognition in adverse conditions: A review," Lang. Cognit. Proc. **27**, 953–978.

Miyake, A. & Friedman, N. P. (**2012**). "The nature and organization of individual differences in executive functions: Four general conclusions," Curr. Dir. Psychol. Sci. **21**(1), 8–14.

Moreno, S. (**2009**). "Can music influence language and cognition," Cont. Music Rev. **28**(3), 329–345.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., and Chau, T. (**2011**). "Short-term music training enhances verbal intelligence and executive function," Psychol. Sci. **22**(11), 1425–1433.

Musacchia, G., Strait, D., and Kraus, N. (**2008**). "Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians," Hear. Res. **241**(1–2), 34–42.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**(2), 1085–1099.

Obleser, J., Wise, R. J. S., Dresner, M. A., and Scott, S. K. (**2007**). "Functional integration across brain regions improves speech perception under adverse listening conditions," J. Neurosci. **27**(9), 2283–2289.

Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (**2009**). "Musician enhancement for speech-in-noise," Ear Hear. **30**, 653–661.

Parbery-Clark, A., Strait, D., Anderson, S., Hittner, E., and Kraus, N. (**2011**). "Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise," PLoS One **6**(5), e18082.

Patel, A. D. (**2008**). *Music, Language, and the Brain* (Oxford University Press, New York).

Patel, A. D. (**2011**). "Why would musical training benefit the neural encoding of speech? The OPERA hypothesis," Front. Psychol. **2**, 142.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (**1998**). "Processing syntactic relations in language and music: An event-related potential study," J. Cognit. Neurosci. **10**(6), 717–733.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.

Rabbitt, P. (**1968**). "Channel capacity, intelligibility and immediate memory," Q. J. Exp. Psychol. **20**, 241–248.

Rönnberg, J., Rudner, M., Lunner, T., and Zekveld, A. A. (**2010**). "When cognition kicks in: Working memory and speech understanding in noise," Noise Health **12**(49), 263–269.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). "Statistical learning by 8-month-old infants," Science **274**, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (**1999**). "Statistical learning of tone sequences by human infants and adults," Cognition **70**, 27–52.

Soncini, F., and Costa, M. J. (**2006**). "The effect of musical practice on speech recognition in quiet and noisy situations," Pro Fono **18**(2), 161–170.

Strait, D., and Kraus, N. (**2011a**). "Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise," Front. Psychol. **2**, 113.

Strait, D., and Kraus, N. (**2011b**). "Playing music for a smarter ear: Cognitive, perceptual and neurobiological evidence," Music Percept. **29**(2), 133–146.

Strait, D., Kraus, N., Parbery-Clark, A., and Ashley, R. (**2010**). "Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance," Hear. Res. **261**, 22–29.

Strait, D., Parbery-Clark, A., Hittner, E., and Kraus, N. (**2012**). "Musical training during early childhood enhances the neural encoding of speech in noise," Brain Lang. **123**, 191–201.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., Tzeng, O., Arévalo, A., Vargha, A., Butler, A. C., Buffington, R., and Bates, E. (**2004**). "A new on-line resource for psycholinguistic studies," J. Mem. Lang. **51**, 247–250.

Watson, C. S., Qiu, W. W., Chamberlain, M. M., and Li, X. (**1996**). "Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition," J. Acoust. Soc. Am. **100**(2), 1153–1162.

Wechsler, D. (**2011**). *Wechsler Abbreviated Scale of Intelligence–Second Edition* (Pearson, Bloomington, MN).

Zekveld, A. A., George, E. L. J., Kramer, S. E., Goverts, S. T., and Houtgast, T. (**2007**). "The development of the text reception threshold test: A visual analogue of the speech reception threshold test," J. Speech Lang. Hear. Res. **50**, 576–584.

Zendel, B. R., and Alain, C. (**2012**). "Musicians experience less age-related decline in central auditory processing," Psychol. Aging **27**(2), 410–417.

Zuk, J., Benjamin, C., Kenyon, A., and Gaab, N. (**2014**). "Behavioral and neural correlates of executive functioning in musicians and non-musicians," PLoS One **9**(6), e99868.